International Journal of Exploring Emerging Trends in Engineering (IJEETE)
Vol. 03, Issue 02, Mar-Apr, 2016 Pg. 165 - 169 WWW.IJEETE.COM

OBJECT RECOGNITION: A SURVEY

S.S.Deore[1], C.V.Longani[2]

Computer Dept., SRES COE, Savitribai Phule Pune University, India

Abstract : Object recognition is concerned with determining the identity of an object being observed in the image from a set of known labels. Oftentimes, it is assumed that the object being observed has been detected or there is a single object in the image. Object Recognition is the critical task in many computer vision applications such as video surveillance, vehicle parking system, person identification, and behavior analysis. Object Recognition area especially for human and vehicle is currently most active research topic. Typically it includes the phases like pre-processing, Background extraction, object detection, recognition and classification. In this paper, we review current methods for the object recognition.

Keywords: object recognition, feature extraction, and object labeling

I. Introduction:

Object recognition is one of the most interesting abilities that humans easily retain since childhood. With a simple scan of an object, humans are able to identity or category even with the appearance variation due to change in pose, illumination, texture, deformation, and under occlusion. The common preparation of the problem is essentially: given some knowledge of how certain objects may appear, plus an image of a scene containing those objects, find which objects are present in the scene and where. Recognition is accomplished by matching features of an image and model of an object. The two most important issues that a method must point out are the definition of a feature, and how the matching is found. Nevertheless, it is a daunting task to develop vision systems that match the intellectual capabilities of human beings, or systems that are able to tell the specific identity of an object being observed. The main reasons can be the factors like: relative position of an object to a camera, lighting variation, and difficulty in generalizing across objects from a set of exemplar images. Important key point of object recognition system is how the regularities of images which are taken under different lighting and pose conditions are extracted and recognized.

In different way we can say that, all the algorithms have certain representations or models to capture these characteristics, thereby facilitating procedures to tell their identities. Historically, two main trends can be identified. In the geometry-or model-based object recognition, the knowledge of an object presence is provided by the user as an explicit CAD-like model. On the other end of the spectrum are the appearance-based methods, where no explicit user- provided model is required. The object representations are usually developed through an automatic learning phase (not necessarily), and the model typically relies on surface reflectance properties. Recently, methods which put local image patches into correspondence are emerged. Models are learned automatically, objects are characterized by appearance of small local features. A global feature of the representation is controlled by weak or strong geometric model.

II. Classes of object recognition methods

There are two classes of recognition methods that are : Generative and Discriminative models. Formally, these categories are described as follows: Given an input x and a label y then a generative classifier learns a model of the joint probability p (x,y) and classifies using p(y|x), which is obtained by using Bayes' rule. In contrast, a discriminative classifier models the

posterior p(y|x) directly from the data or learns a map from input to labels: y=f(x).

Generally in recognition system, giving the training data and the corresponding labels the goal is to find optimum decision boundaries. Thus, to classify an unknown sample using a discriminative model a label is assigned directly based on the estimated decision boundary. In contrast, for a generative model the chance of the sample which is estimated and the sample which is assigned may belongs to the most likely class. This paper mainly focuses on generative methods, i.e., the goal is to represent the image data in a suitable way. Therefore, objects can be described by different signs. These include model-based approach, shape-based approach and appearance-based models. Model-based approaches try to characterize (approximate) the object as a collection of three dimensional, geometrical primitives like boxes, spheres, cones, cylinders, generalized cylinders, surface of revolution, whereas shape-based methods represent an object by its shape/contour. As compared to this, appearance-based models focus only on the appearance of the object, which is usually captured by different two-dimensional views of the object-of-interest. Based on the applied features these methods can be sub-divided into two main classes, i.e., local and global approaches [1].

A. Geometry-based or Model- based approach:

In paper [2] author briefly discussed about this approach, they stated that, Model-based object recognition algorithms are based on comparatively simple CAD wire models of objects, as illustrated in Figure 1. Using such models, the starting and end points of lines can be estimated very efficiently into the image plane, allowing real-time tracking of objects with relatively low computational effort.

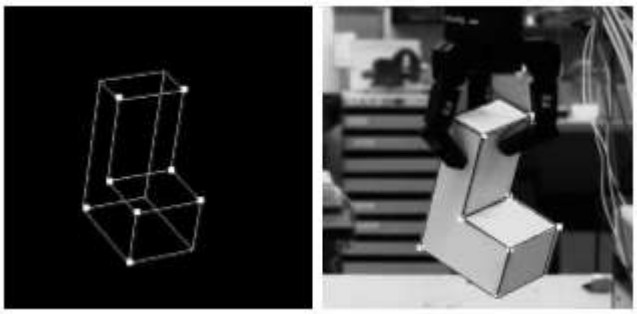

Figure1. Illustration of an object modeled by a wire model

However, the limits of such systems are nothing but the shapes to which they deal. Most real-world objects, such as cups, plates and bottles cannot be represented in this manner. As shown in Figure 2, where the Southern Cross becomes clear when taking a look at an object with a complex shape.

Figure 2 Illustration of a 3D model of a can

In paper [4] author stated that, this approach requires two representations: one to represent object model, and another to represent the content of the image. To facilitate finding a match between model and image, the two representations should be related closely. In the idyllic case there will be a simple relation between primitives used to describe the model and those used to describe the image. Would the object be, for example, described by a wire frame model, the image might be best described in terms of linear intensity edges. Each edge can be then matched directly to one of the model wires. However, the model and image representations often have distinctly different ”meanings”. The model may

describe the 3D shape of an object while the image edges correspond only to visible manifestations of that shape mixed together with ”false” edges (discontinuities in surface) and illumination effects (shadows)[2].

The main disadvantages of geometry-based methods are: the dependency on reliable extraction of geometric primitives (lines, circles, etc.), the ambiguity in interpretation of the detected primitives (presence of primitives that are not modeled), the restricted modeling capabilities only to a class of objects which are composed of few easily detectable elements, and the need to create the models manually.

B. Appearance-based algorithms

The appearance of an object is the combined effect of its shape, reflectance properties, poses in the scene, and the illumination conditions. While shape and reflectance are intrinsic properties that do not change for any rigid object, pose and illumination vary from one scene to the next.

In this method, firstly the model is constructed from the dataset image which includes different orientations, different illuminations and multiple instances of a class of objects. Then in next stage the query image segmented by texture, color or motion. Then recognition system compares an extracted part of the query image with the dataset images. Appearance based algorithm widely classified in two approached as local and global approach.

1. Local appearance based approaches: It recognize and localize objects on the base of local features. The use of local features always depends on extracting textural information. In this method objects are represented by a set of local features, which are automatically computed from the training images. When a query image is given, local features are extracted as in the training images. Similar features are then retrieved from the database and the presence of objects is assessed in the terms of the number of local correspondences. Since it is not required that all local features match, the approaches are robust to occlusion and cluttered background. To recognize objects from different views, it is necessary to handle all variations in object appearance. The variations might be complex in general, but at the scale of the local features they can be modeled by simple, e.g. affine, transformations. The detailed description of this method is given in paper[3].

Several methods have been proposed for feature extraction, among which the most popular are the Harris corner detector [6], Shi-Tomasi features [7], SIFT features [8], and Maximally Stable Extremal Regions [9]. All object recognition and localization systems based on such features depend on the successful extraction of a sufficient number of features for each object this is only possible for objects containing enough features which is not the case for many objects in our kitchen environment. For example, only rarely dishes contain real texture features, since they are often solid colored with only very few textural information. For such objects, it is more sensible to assume the objects can be segmented, e.g. by color, and solve the problem of recognition and localization with a global appearance-based approach

2. Global appearance – based approach:

In this approach for each object, a set of segmented views are stored, covering the space of possible views of one object. By associating pose information with each view, it is possible to recover the pose through the matched view from the database. For reasons of computational efficiency, PCA [5] is applied for reducing dimensionality. Now a day, the trend goes toward local appearance-based methods using texture features. Selection of any one approach is depends on the application. In paper [3] author used this method for real time object recognition.

Limitation of Appearances based approach is, they required isolation of the complete object of interest from the background. Thus they are sensitive to occlusion and require good segmentation. A segmentation method includes global histogram matching methods, color histogram. Now days the concept of histogram matching is generalized by responses of various

filters to form the histograms (called then receptive field histograms).

C. Shaped based Approach:

In this approach, we require to extract the silhouette of the moving object in each video frames then create a set of representative model views organized in a view graph ; and match the object and model silhouettes based on their shape while maintaining motion coherence over time as shown in figure 3. This approach relies on good motion segmentation, which in our case is achieved by jointly segmenting the video frames.

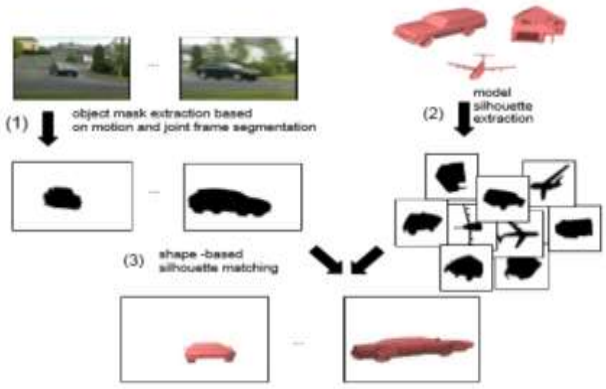

Figure 3. Recognition in videos by matching the shapes of object silhouettes obtained using motion segmentation with silhouettes

This method is used in system discussed in paper[10], in which they have extracted the object silhouettes which fuses two processes (see figure (4)): (1)feature tracking and motion-based clustering of the resulting tracks as either object or background; (2) video segmentation into region tracks which represent parts of the scene evolving over time. In a subsequent step, they combined the feature track labeling with the segment tracks to obtain masks for the object in each frame. Purpose of this technique is to achieve robust sparse motion segmentation, while region tracks will propagate this motion segmentation to the whole image, and thus the object silhouette can be extracted. Then shaped based matching is performed on the query imaged as discussed in paper[10].

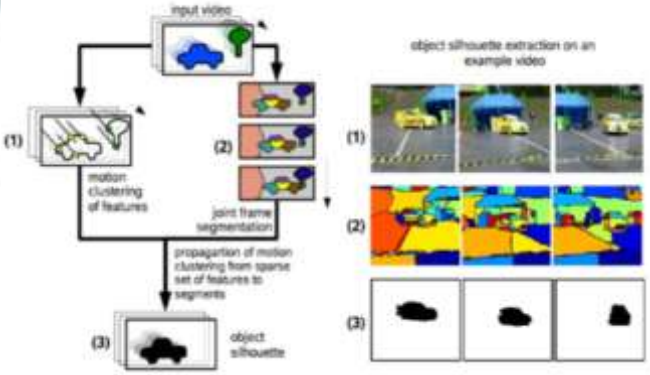

Figure 4 On the left is a schematic of the object silhouette extraction, and on the right is an example on a car video: (1) feature clustering based on common motion; (2) segment tracks; (3) object silhouette.

III. Conclusion

We have surveyed the methods involved in recognition of objects from a image and videos. The advantages and disadvantages of each of these methods were discussed. In the future, we plan to propose an algorithm that overcomes the disadvantages of the existing object recognition methods and classify objects in a video and image with capability to handle multiple objects and occlusion.

IV.References

1.Peter M. Roth and martin winter," survey of appearance-based Methods for object recognition",Inst. for Computer Graphics and Vision Graz University of Technology, Austria

2. Pedram Azad, Tamim Asfour and Ruediger Dillmann,"Combining Appearance-based and Model-based Methods for Real-Time Object Recognition and 6D Localization",Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems October 9 - 15, 2006, Beijing, China

3. Jırı matas and stephenn obdrzalek,"Object recognition methods based on transformation covariant features"Center for Machine Perception, Czech Technical University, Prague, 120 35, CZ
4.Dunteman,Principal Component Analysis. Sage Publications, 1989

5.Bastian Leibe and Bernt Schiele,"Analyzing Appearance and Contour Based Methods for Object Categorization",Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03), Madison, USA, June 2003.

6.Harris and M. Stephens, "A combined corner and edge detector," in Alvey Vision Conference, Manchester, UK, 1988, pp. 147–151.
7. J. Shi and C. Tomasi, "Good features to track," in International Conference on Computer Vision and Pattern Recognition (CVPR), Seattle,USA, 1994, pp. 593–600.

8. D. G. Lowe, "Object recognition from local scale-invariant features," in International Conference on Computer Vision (ICCV), Corfu, Greece,1999, pp. 1150–1517.

9. S. Obdrzalek and J. Matas, "Object recognition using local affine frames on distinguished regions," in British Machine Vision Conference(BMVC), vol. 1, Cardiff, UK, 2002, pp. 113–122.

10. Alexander Toshev etc.al. ,"Shape-based Object recognition in Videos Using 3D Synthetic Object Models "

